DOCUMENT RESUME

ED 363 643 TM 020 664

AUTHOR	Hymes, Donald L.
TITLE	Making Sense of Testing & Assessment.
INSTITUTION	American Association of School Administrators, Arlington, Va.
REPORT NO	ISBN-0-87652-190-1
PUB DATE	93
NOTE	32p.
AVAILABLE FROM	American Association of School Administrators, 1801 N. Moore St., Arlington, VA 22209-9988 (AASA Stock No. 21-00399).
PUB TYPE	Guides - Non-Classroom Use (055)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Accountability; Achievement Tests; Aptitude Tests; *Children; Criterion Referenced Tests; *Educational Assessment; Educational Trends; Elementary School Students; Elementary Secondary Education; Intelligence Tests; Norm Referenced Tests; *Parent Participation; Performance; Secondary School Students; *Standardized Tests; Student Evaluation; Test Bias; Test Reliability; *Test Use; Test Validity; Test Wiseness
IDENTIFIERS	*Alternative Assessment; *Performance Based Evaluation

ABSTRACT

Learning about tests will help parents help their children perform better on all types of assessments. This guide explains the many types of tests that students can take and describes the benefits and limitations of each. Test use and improving test performance are discussed. Assessment usually takes place for purposes of instruction, screening, and accountability. While classroom tests are generally created by the teacher, students frequently face standardized tests, whether in the form of achievement tests, aptitude tests, or intelligence tests. Differences between norm-referenced and criterion-referenced tests and their uses are discussed. Issues of test bias and the fair use of tests are considered. Suggestions are given to help children develop test-taking skills. Alternative assessment, particularly performance based assessments, are discussed as a growing trend in American educational assessment. Three parent resources are listed, and a glossary is provided of testing terms. Making sense of testing and assessment is recognized as requiring commitment from parents and schools to work together. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 363 643

MAKING SENSE

Testing & Assessment

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W. Henry

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM020664

American Association of School Administrators

BEST COPY AVAILABLE

Copyright © 1993, American Association of School Administrators
ISBN: 0-87652-190-1
Library of Congress Catalog Card No.: 93-072247
AASA Stock No.: 21-00399

Contents

3 Introduction
- What Should Parents Know?
- What Testing Is All About

5 Types of Assessments
- Classroom Tests
- Standardized Tests
- Norm-Referenced Tests
- Criterion-Referenced Tests

13 Testing and Your Child
- How Can I Help My Child?
- Are All Tests Reliable and Valid?
- Keeping the Playing Field Level

19 Alternative Assessments: The Future is Here
- Performance-Based Assessments
- Portfolios
- A Long Road Ahead

24 Conclusion

25 Glossary

Introduction

"I have a test tomorrow!"

How many times have you heard that from your child? What is your usual response? Do you simply say, "That's nice, good luck," or do you cancel everything for the evening and get out the books?

Either response might be the right one depending on the type of test.

That word "test" covers everything from a pop quiz containing five spelling words to a half-day marathon in the gym with pencils and tight security. Generally, today's students are being tested more than ever before. For example:

- A test may be used to help determine a child's placement in reading.
- Teachers test their students regularly to find out how well each one is learning in class.
- School districts give tests to get an idea about how well all students are doing in math.
- Some state tests might be used to compare school districts.
- A test might be required for a student to be promoted to the next grade level or to determine whether a high school student receives a diploma.
- A student's admission to college may hinge, at least in part, on a test.

For these reasons and more, we need to understand various types of tests and how their results can be—and should not be—used.

What Should Parents Know?

If you're a parent, learning about tests will help you help your child perform better on all types of assessments.

This booklet can help each of us understand:

- The many types of tests students may take.
- What each of these tests is designed to find out.
- The benefits and limitations of each type of test.
- The way test results are used by the teacher, the school, and the school district.
- How to help improve your children's performance on tests.

A glossary is included on page 25. You may want to read over it to familiarize yourself with the testing terms used throughout this booklet.

What Testing is All About

Looking back, many people remember the weekly arithmetic quizzes in elementary school and the final exams each semester in high school. And then there were the special tests—the ones in the gym or cafeteria—where students battled the clock as they filled in blanks on an answer sheet.

Those tests are still around, along with many new ones. In fact, schools use many different methods to evaluate a child's work. As a result, most educators prefer the word assessment to define the process of evaluating student progress. Tests in the old-fashioned sense are only part of this process, but they still are very important.

Although some tests are mandated by state boards of education, few school systems in the nation have identical assessment programs. The best way to find out what tests your children will be taking is to ask their teachers or counselors.

As you begin to learn about tests, you'll probably be confronted with many unfamiliar words and acronyms—terms like standardized, norm- and criterion-referenced, percentile, aptitude, performance-based, IQ, SAT, and ITBS. They may be puzzling at first, but with the help of the school staff—and this booklet—you can better understand their uses and how these assessments are important milestones in your child's education.

Why test?

Assessments take place for three basic reasons:

Instruction. Assessments may be used to determine how well a child is understanding classroom lessons and developing skills. Information gathered from these tests is used by the teacher to improve student learning in the classroom. Or, administrators and curriculum directors may use assessments to make informed decisions on what is taught.

Screening. Various tests give teachers and educators an idea if students are ready to start school, ready for more advanced work, or, on the other hand, if they need particular kinds of remedial help.

Accountability. The public wants to know how its investment in education is paying off, especially during times of economic hardship. Educators, school boards, districts, states, and various other groups are frequently required to "prove" how well students are doing. Rankings on state or national tests often are used to illustrate school performance.

Types of Assessments

Classroom Tests

These assessments are created by the teacher and are based directly on what has been taught in the classroom. They can range from five-minute pop quizzes to multiday, year-end tests. They can be multiple-choice or fill-in-the-blank tests, book reports, oral presentations, essays, or actual demonstrations of certain skills (such as a science experiment). And classroom assessments don't even have to be tests. Classroom observations are an important part of a teacher's assessment process.

Classroom assessments have these purposes:

- To review how well students have learned specific lessons.
- To determine each student's strengths and weaknesses so that instruction can be directed to areas that need further attention.
- To determine each student's report card grade.

What is important to understand is that a classroom test often is unique to a particular teacher and class. The results are not meant to be compared with students' work in other classes.

Standardized Tests

On the other hand, standardized tests are given to large numbers of students in different classrooms and schools. The tests are given under similar conditions to students at the same grade level and at approximately the same time during the school year.

The assumption behind standardized testing is that if everybody gets the same test and the same instructions and has the same amount of time to finish, comparisons of test results will be meaningful and, thus, can be used to make many educational and administrative decisions.

Depending on the purpose of the test and how it is scored, a standardized test can show:

- Whether a student is ready for a certain level of instruction.
- Whether a student has mastered learning objectives set by the school system or the state.
- How one student stands in relation to his or her classmates.
- How one class is doing in various subjects compared with other classes in the school.
- How one school compares with other schools throughout the district, state, or country, in specific subjects.
- How students in various states or regions compare to each other.

Thought-Provoking Questions

Remember taking standardized tests of multiple-choice and fill-in-the blank questions that tested your knowledge of mathematics or historical events?

1. 3.45 + 8.5 + 10.7 =

a. 21.57 **b.** 22.92 **c.** 22.65 **d.** 21.12

2. Magellan was the first person to:

a. go around the world **b.** discover the Atlantic Ocean **c.** land on American soil **d.** look for the Fountain of Youth.

3. George Washington was born in the year _____.

Then there were analogies to test your ability to form relationships among different groups:

4. Governor is to State as President is to:

a. the Cabinet **b.** the Senate
c. the United States **d.** the Electoral College

However, today's standardized test questions have become more complex in an effort to test students' thinking skills, and not just their memories. These math examples require students to problem solve and reason how to arrive at the correct answer:

5. Each of the following two-digit numbers has one digit covered. Which of these five numbers is the only possible multiple of 12?

a. _9 **b.** _5 **c.** _3 **d.** 3_ **e.** 5_

6. Two married couples, the Mercados and the Smiths, are seated in the first four seats of a theater row. If Mr. Mercado sits in the aisle seat, how many arrangements of the other three people are possible so that no person is sitting next to his or her spouse?

a. 2 **b.** 3 **c.** 4 **d.** 5

7. A company will use vans to carry employees to a company picnic. Each van can carry 8 people. How many vans are needed to carry 36 employees to the picnic?

a. 3 **b.** 4 **c.** 5 **d.** 6 **e.** 8

8

Standardized tests take many forms

Here are some brief descriptions of the different kinds of standardized tests a student might be expected to take at some point during a school career.

Achievement tests. By far, most of the tests given in school, like all classroom assessments, are "achievement tests." They are used to measure a child's school-taught learning.

The most common standardized achievement tests are the:

- Iowa Test of Basic Skills (ITBS)
- California Achievement Test (CAT)
- Metropolitan Achievement Test (MAT or MET)
- Comprehensive Test of Basic Skills (CTBS)
- SRA (Science Research Associates) Achievement Series.

These tests are developed by commercial test publishers and marketed to school systems nationwide. They are written for selected grade levels and are routinely administered every year to measure students' achievement in selected subjects, such as reading, math, social studies, and science.

In addition, many states are developing their own achievement tests based more directly on the local curriculum.

Aptitude tests. In contrast to achievement tests, aptitude tests are given before instruction begins and are used to determine a student's potential to do certain work. The questions explore students' familiarity with words and mathematical concepts and their ability to think through problems.

Most aptitude tests are used to screen students to make decisions on placement, such as who should be enrolled in advanced placement or gifted classes.

The best-known aptitude tests are those that are supposed to predict how well high school students will do in college. These are the Scholastic Aptitude Test (SAT)* and the American College Test (ACT), taken by more than 2 million high school juniors and seniors every year. They are intended to measure only a student's ability to do college-level work. These tests do not pretend to assess general student achievement in high school, and they should not be used to evaluate a high school's performance or to compare schools.

Intelligence tests. These tests have been around for at least 100 years, and at the peak of their popularity were used to measure young children's innate "intelligence," represented by

** The College Board recently changed the name to Scholastic Assessment Test.*

The Nation's Report Card

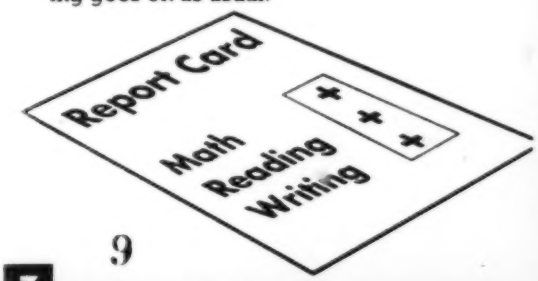
Perhaps you've seen alarming newspaper headlines warning that America's students are going downhill in math, reading, or writing. These stories probably are referring to results from the Nation's Report Card, or National Assessment of Educational Progress. NAEP is a federally funded program of tests intended to measure, over time, what American students know and can do in various subject areas, including math, reading, science, writing, history, civics, and other areas. These assessments have been given to a sample of students at ages 9, 13, and 17 since 1969.

Because of the cost of administering NAEP tests to a larger number of students, as well as other practical limitations, standardized NAEP assessments are given only to a small sample of the country's student population. NAEP results form the basis for various government reports, but they have limited practical value for parents or local educators.

In 1990, however, a NAEP-issued report sparked a controversy for decrying a lack of improvement by the nation's schoolchildren over the last 20 years. Based on this report, some federal officials proposed a national test and standards so that policymakers and parents could keep an eye on the "big picture."

A national coalition of teachers, administrators, and parent groups promptly challenged national standards and testing, arguing that federal involvement in testing and especially in setting standards would endanger the traditional concept of local control of education.

A federal commission was formed in 1990 to begin the difficult task of considering national standards. Meanwhile, NAEP testing goes on as usual.



Controversy over SAT Scores

Many critics have cited a decline in average SAT scores since 1965 as further proof that America's schools are failing to provide students with the education they need to succeed in life. However, this argument doesn't consider that each year increasing numbers of students are taking the SAT and that these students come from diverse backgrounds. In the past, mostly high-achieving students from wealthy families took the test because they were the students most likely going on to college.

During the 1989-90 school year, more than 1 million high school students took the test. Naturally, a large influx of new students taking the test will cause a decline in average SAT scores. But this decline is not completely negative. On the plus side, more students than ever before are considering some form of higher education after high school.

an "IQ," or intelligence quotient. The IQ became a permanent entry on each school record, and in some cases, was used to assign students to "tracks."

In recent decades, however, educators are less likely to make widespread use of intelligence tests. Many believe a child's IQ is affected by outside influences and can change throughout life. IQ tests are most often used today after a child is referred to a psychologist for a diagnosis of school problems.

Interest and aptitude tests. Usually given toward the end of junior high school, interest and aptitude tests help students and their parents decide what type of courses the student would like to take or might do well in during high school. These tests also are used as a preliminary measure of career interests.

Competency tests. These include a variety of standardized achievement examinations used to verify that a student has reached a specified level in a subject.

Such tests, which also may be called "functional," "mastery," or "proficiency" tests, have been developed in several states in response to public demands for accountability.

Competency testing programs are designed to make sure students have met the curriculum objectives set by the state or the school system. Students who fail may be held back or required to take remedial courses until they pass. In some cases, students must pass these tests before they graduate.

The standards set by the state or school system for passing these tests, however, are likely to be at a minimal level. As a result, some educators have criticized this type of testing—which also is called "minimum competency"—because it emphasizes "passing" instead of "excellence." The most recent trend is for state testing programs to measure much higher standards.

Norm-Referenced Tests

For many decades, a key concept in testing accountability has been the norm-referenced test. Norm-referenced tests fall into the broad category of standardized tests. In this book, however, norm-referenced tests have been given a separate section because they're so important. Norm-referenced means that local or district scores are compared against a nationwide sample of students who already have been given the test under the same conditions by the publishers. This sample is called a norm group.

Students taking norm-referenced tests are not graded on the percentage of questions right or wrong, but are ranked according to how well they did in relation to other students.

Since every school that subscribes to one of these tests is ranked according to the same norm, local educators can see where the district stands in relation to the rest of the state or nation. In addition, looking at how test scores change from year to year allows educators to spot trends that should be addressed. For example, many educators use these tests to answer important questions, such as:

- Are scores improving in certain subjects because of the success of recent curriculum efforts?
- Are scores declining in certain subjects, indicating the need for special attention?
- Are specific schools or groups of students showing or not showing expected progress?
- Are minority students improving in relation to majority students?
- Are girls closing the gap with boys in math and science?

High-stakes testing. You've probably read about the results of norm-referenced tests in newspaper headlines. School administrators and elected officials alike monitor these scores

Proficiency Testing in Ohio

To graduate, high school students in Ohio must pass a four-part state proficiency exam in reading, writing, mathematics, and citizenship. A science section will be added in 1995-96.

Starting in the ninth grade, students have at least two opportunities a year throughout high school to pass all four sections. Students do not have to retake any sections they already have passed.

The test contains a writing prompt for students to respond to, as well as multiple-choice questions in the other three subjects. The State Board of Education determined how many questions students must answer correctly and what score they must get on the writing sample to pass each of the tests.

carefully because they may lead to important administrative and political decisions, including school funding, staff ratings, and training needs. Norm-referenced scores also could affect a community's support for its schools. Consequently, norm-referenced tests often are referred to as "high-stakes" tests.

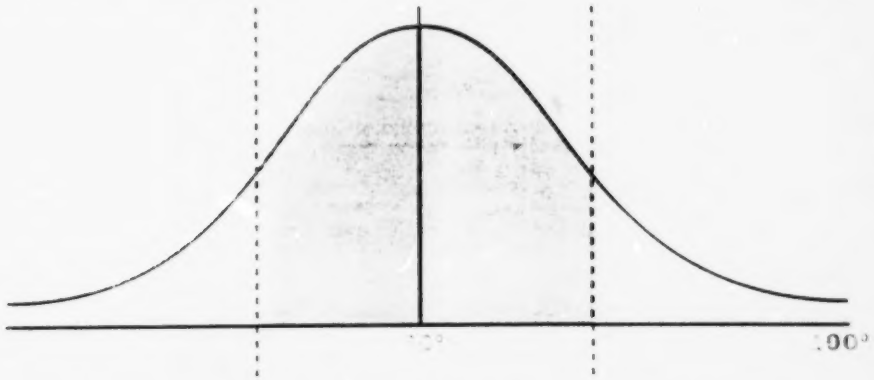
Yet, norm-referenced tests tell educators and parents little about how well students are actually mastering required classroom skills. In fact, some studies have s' own these national tests cover only 25 to 30 percent of what might be covered in a child's class. As a result, when these test scores are stressed, teachers and administrators may feel pressured to concentrate on topics they know are on the test solely to boost scores (see Teaching to the test on page 15).

How norm-referenced tests are reported

When the results of norm-referenced tests are sent home to parents or published in a school newsletter or local newspaper, they may seem to be in a foreign language. This is because when test scores are used for comparisons, the familiar A-B-C grades or percentages of right answers would be meaningless. Instead, what is important is how students' scores stack up with those of other groups.

Norm-referenced test results are reported in a variety of ways, all of which compare each child's score with the average of other students' scores or with those of a norm group.

The bell curve. In norm-referenced tests, scoring begins by constructing what is called a "normal curve," or because of its distinct shape, a "bell curve."



The shape of the curve determines what percentage of the scores falls into each of its segments. The curve is always balanced, with 50 percent of the norm group above the midpoint and 50 percent below.

As the illustration shows, only a small number of students falls at each extreme, while a large percentage falls in the middle, or "average" range.

Some critics argue the bell curve encourages mediocrity because its goal is to have most students clustered around the "average" range of the curve.

Grade-equivalent scores. These are reported by two numbers separated by a decimal point. The first stands for a grade in school, and the second for a tenth of the school year. A grade equivalent score of 3.5 means that a student scored as well on that test as the average student halfway through the third grade. Thus, if a third-grader achieves a 3.8 on a test given in November, he or she can be considered about five months ahead of the average pace of 3.3.

Percentile scores. Contrary to their name, percentile scores have nothing to do with the percentage of questions the student got right. Percentiles, with a low of 1 and a high of 99, show how students compare with the norm group. A percentile score is the percentage of students your child scored as well as or better than. If a student scores at the 65th percentile, it means that he or she performed as well as or better on the test than 65 percent of the norm group.

The 50th percentile is considered average, and the largest percentage of students is clustered between the 40th and 60th percentiles.

Quartiles or quintiles. After the percentiles are determined, the scores may be grouped in other ways for reporting purposes, depending on the school system's preference.

When the percentiles are grouped into four categories, they are called quartiles. Students who scored in the 1st through 25th percentile are in the first quartile, those between the 26th and 50th are in the second quartile, and so forth. When the percentiles are arranged into five groups (0-20, 21-40, and so on), they are called quintiles.

Stanines. Like percentiles, stanines rank students, but in nine groups. Stanine scores of 1, 2, and 3 are considered below average; 4, 5, and 6 are in the average range; and 7, 8, and 9 are above average.

Criterion-Referenced Tests

A criterion-referenced assessment tests what students actually have been taught. They are designed to measure how much knowledge a student has gained from instruction.

For example, the classroom test a teacher develops on what students should have learned is a criterion-referenced test. Rather than ranking students with a norm group, the classroom test measures how many of the objectives each student has mastered. Thus, a grade of 75 on a classroom test means that a child has mastered three-quarters of the learning objectives.

But the classroom test is one-of-a-kind. To measure overall school or district performance on learning objectives, a standardized criterion-referenced test is needed—one that is given in the same form under the same conditions to large numbers of students. These tests can show a teacher and a parent how well an individual student can do long division at the same time they show the principal and superintendent how students are meeting the district's specific curriculum objectives in math.

Local control. Realizing the limitations of national tests in measuring actual learning, many school districts and state departments of education are developing their own standardized, criterion-referenced testing programs. These tests are "standardized" because they are administered to many students at the same time and under similar conditions. They are criterion-referenced because they are based on local learning objectives.

Today, more and more schools are replacing norm-referenced tests with criterion-referenced tests because they provide better information about what a student actually has learned.

Unlike traditional paper-and-pencil, multiple-choice tests, a growing number of criterion-referenced tests are "performance based"—meaning that the student must demonstrate what he or she has learned, not just fill in a bubble on an answer sheet.

Standardized criterion-referenced test scores usually are reported in percentages, which, in this case, do represent the number of questions the student answered correctly. Unlike norm-referenced tests that aim to have most students score in a middle-of-the-road, "average" range, the goal of criterion-referenced tests is to have as many students as possible score at the upper end of the scale, because it means they are learning the material being covered.

Testing and Your Child

When your child comes home with a score of 90 on a weekly, 10-question math quiz, you know what that means: He or she got 9 out of 10 questions right. If you look further, you may see that the missed question dealt with multiplying fractions. The teacher will see that too, and will likely give your child a little extra work or coaching on fractions next week.

However, interpreting a norm-referenced test is not that simple.

Let's say the school sends home results of the latest Iowa Test of Basic Skills, reported in percentiles. Your third-grade student gets a 68 on the verbal test and an 82 on math.

First, you should be pleased that your child scored above average in both areas. You will, no doubt, be concerned about the disparity in scores between subjects. You might want to schedule a conference with the language arts or reading teacher to discuss what verbal areas your child might be struggling with and why. Perhaps your child also would benefit from additional practice in certain verbal skills.

Look at the big picture. As a parent, you should not underestimate the importance of standardized tests, but also understand that these scores represent only a small part of your child's actual school performance. Overall, the best you can say about the results of national standardized tests is that they provide a very generalized picture of school performance.

Misunderstanding test results and jumping to false conclusions is the biggest risk of standardized testing. That's why it is important to understand the difference, for example, between a norm-referenced test and a competency test, and to recognize when a particular score should be a cause for concern.

How Can I Help My Child?

Parents should be involved in their children's education throughout the school year, not just at test time. In this way, the tests will be seen as just one tool your school uses as part of its instructional program.

The best advice is to maintain close contact with your children's teachers, counselors, and principals. Ask for explanations about the various types of tests they give. Here are some suggestions on ways you can help ensure that your children do their best on tests:

- If you know what tests are scheduled and what their purposes are, you and your children will be prepared even if the school doesn't send a special notice home.
- Have a serious talk with your children about each test. Let them know how interested you are in the tests and how important it is that they do their best.

- Explain what will be done with the results. As alarming as it sounds, many students regard tests, particularly standardized ones, as a nuisance and, in some cases, deliberately mark wrong answers in protest.Or, they may randomly mark answers just to get finished. Depending on the purpose of the test, this indifference could have a long-term impact on the child's school career.
- If materials are available from the school, spend time with your children reviewing the subject matter and test format.
- Try to shelter your children from distractions before the test.
- Make sure your children get a good night's sleep before the test and eat a healthy breakfast that morning.
- Get your children to the school or bus with plenty of time to spare.
- If allowed, let your children bring their own materials (pencils, pens, calculators, protractors) to the test to help them feel comfortable.

Seven Questions To Ask Your Child's Teacher

Here are some questions to ask, starting early in the school year:

1. Is there a testing calendar for the entire year?

2. Are there things I can do before certain tests are given to help my child do better? Are there study tools or classes available to help my child prepare?

3. What are the purposes of each test?

4. What are the consequences of each test? Who gets to see the results? Do they become part of my child's permanent record?

5. Can my child retake any of the tests if the results are unsatisfactory?

6. Is there anything I can or should do as a result of my child's performance on a test? Is there anything I should not do?

7. How do the results of each test fit with the bigger picture of my child's performance in school?

The answers to the above questions will help you play a more valuable role in your children's schooling.

- When the results are announced, make it a point, if you have questions, to ask the teacher or counselor about the school's scores, as well as your children's. Make sure you and your children understand what they mean.
- So that you don't place undue emphasis on a single test, stay involved with the school throughout the year. That will give you a clearer picture of your children's total school progress and areas that need improvement.
- Always encourage your children to do their best, constantly emphasizing how important education is to you. But at the same time, don't apply too much pressure by setting unrealistic goals. Creating test anxiety can easily backfire.

Are All Tests Reliable and Valid?

In the language of assessment, a test is considered reliable if it yields consistent and similar scores for students of like abilities or knowledge. But many factors other than a child's knowledge or ability can have a marked impact on the final score. For example, a child might have been hungry, cold, upset, sick, or distracted during a test. Or, he or she might have been absent the day information on the test was taught.

While test publishers try hard to produce reliable tests, it is obvious that the score only represents a student's performance on a specific sample of questions at a specific point in time.

But even if a test turns out to be highly reliable, the results must still be interpreted with caution, because there is also the question of its validity, which is really more important.

An assessment is considered valid when it really tests what it is supposed to test.

Keeping the Playing Field Level

By definition, a standardized test is given to similar groups of students under similar conditions. But what about situations that give certain students an unfair advantage—or disadvantage?

Teaching to the test

Ideally, the driving force behind instruction should be the curriculum, but when test scores become all-important, instruction may become focused on what the test writers—not what the schools or local community—think children should know.

If raising scores becomes the sole reason for testing, some teachers or schools are tempted to resort to the questionable practice of teaching to the test, in which the teacher can:

- Spend excessive time on topics they know will be tested instead of on the required curriculum.
- Emphasize drill and practice rather than essays to prepare students for the multiple-choice format.
- Have students memorize factual tidbits instead of challenging them to develop their thinking skills.

But does increasing the scores in this manner improve student achievement? Not necessarily.

The Question of Bias

No issue is more sensitive in the testing arena today than that of bias. Especially in areas where minority students score lower than their majority counterparts, critics charge that test questions discriminate against some children because of race, gender, handicapping conditions, geographic location, socioeconomic status, or national origin.

Bias in standardized tests is a difficult problem to pinpoint. Standardized tests are written by educated individuals who live in a particular part of the country. Possibly, their backgrounds will be unintentionally reflected in the language and context of the questions.

The stylized English common to the traditional tests could present quite a barrier to children whose English is "nonstandard," such as refugees from Asia and Latin America, as well as many U.S.-born African Americans.

Other obstacles can surface as well, like references to situations children cannot relate to. A child raised on a family farm may have problems grasping concepts about subways or skyscrapers, as would an inner-city child questioned about farm animals.

Because some words mean different things in different parts of the country, children from one region could interpret a question in a way not intended by the writer of a test. Simple misunderstandings between the words "soda" and "pop," or "pail" and "bucket," could confuse a child. A "hill" in one part of the country is a "mountain" in another. And so on.

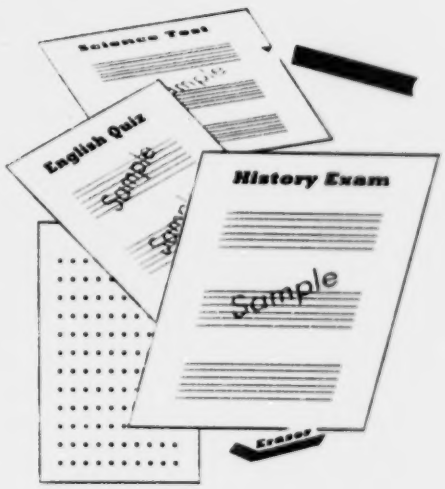
In *How Schools Shortchange Girls,* a 1992 report on gender bias in education, the American Association of University Women said standardized tests are biased against women "when the number of references to or characters of one sex exceed the numbers represented by the other sex or if roles are presented stereotypically." For example, an analysis of the reading comprehension passages in the 1984-85 SAT found references to 42 men and only three women. Thirty-four of the male references cited famous men and their work. Only one of the female references cited a famous woman, and her work was criticized.

In responses to charges of bias, many testmakers have put members of various ethnic and racial groups on review committees to help root out discriminatory questions. But the subtle nature of testing bias makes it a difficult problem to correct. While progress has been made, testing bias is an area that still bears watching.

Test-taking skills

On the other hand, it wouldn't make sense to test children by ushering them into an unfamiliar room, giving them pencils, handing out papers they've never seen, and expecting them to realize their potential.

For this reason, publishers distribute sample test forms and answer sheets so that students can become familiar with the process. Also, many schools and commercial outfits offer courses in test-taking skills, giving students tips on how to approach this important venture. These courses concentrate on increasing students' comfort level by exposing them to the test format, the type of answer sheet to be used, and the way they are supposed to record their answers.



Before the test, students also will find it helpful to know:

- How to eliminate some of the multiple-choice answers to the point where an educated guess is better than leaving a blank.
- If a specific test penalizes wrong answers, which gives them an indication of when it is wise to guess and when it isn't.
- How to conserve time, such as finishing all the questions they know and then going back to work on those they don't.

Worksheet

Questions I'd like to ask my children's teachers:

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____

20

Alternative Assessments: The Future Is Here

The entire field of assessment began to change dramatically in the early 1990s. In place of traditional tests, many states and school systems turned to other ways of measuring student performance. These were called alternative assessments.

Performance-Based Assessments

Instead of picking answers from a ready-made list, performance-based assessments require students to demonstrate that they know an answer. The emphasis is on thinking, not just on mechanical computations. For example, students may be asked to:

- explain historical events
- generate scientific hypotheses
- converse in a foreign language
- conduct research on an assigned topic.

Even in mathematics, it isn't enough just to get the right answer. With a performance-based assessment, the student has to explain how he or she arrived at that answer.

A growing trend. Performance-based assessments are gaining popularity across the country. Vermont, Kentucky, Arizona, and Maryland already operate alternative testing programs.

Washington, Indiana, and Pennsylvania are gradually phasing in performance-based methods to replace or complement their standardized testing programs. In the next few years, other states are sure to follow their lead.

Testing Trends Across America

■ Phasing in Performance-Based Testing

■ Performance-Based Testing already in use

Here are some examples of methods educators frequently use for performance-based assessments:

Open-ended questions

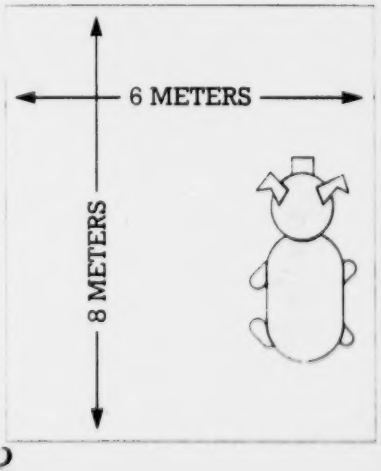
Open-ended questions require students to explore a topic in detail, either orally or in writing. Rather than just choosing an answer, the student must arrive at one and explain how he or she got there. A high school mathematics teacher might ask students to solve the following question, describing each computation along the way to the final answer:

> A room has 8 tables that measure 3 feet by 6 feet each. The tables are made so that they have to be fastened together in one of three ways:
>
>
>
> When all of the 8 tables are fastened together, how could they be arranged to give maximum seating? Draw a sketch and explain why you think this arrangement gives the most seating.

In other subjects, students may be asked to analyze a poem or other literary work or to present arguments a historic character would make about a particular idea.

Not just for high school students. Open-ended questions can be used at any grade level. This third-grade performance assessment in math is based on the popular book, *Charlotte's Web*.

> Wilbur is really a special pig. Mr. Zuckerman let him stay in a nice pen inside the barn. He also had a yard outside the barn that was the shape of a rectangle. Wilbur's yard was 8 meters long and 6 meters wide. Make a drawing of the yard. Label the length of all 4 sides.



22

One day Wilbur needed to walk and think. If he started at one corner of his yard and walked around all 4 sides of the yard, how far did he walk? If he went around the yard 3 times, how far did he walk?

Write a sentence to explain how you got your answer.

Think and write. Remember how history tests always seemed to revolve around dates and explorers? To answer these open-ended history test questions, students have to go beyond rote memorization and understand the differing ideas and complexities of each topic:

1. What if the United States had decided not to drop the atomic bomb?
2. How might our lives be different if America had lost the Revolutionary War to Great Britain?
3. What if Japan had been the first country to colonize the west coast of North America?

Projects

In this type of performance-based assessment, several students may combine forces to solve a problem posed by the teacher, such as a science experiment in a class laboratory. However, English, social studies, and many other subjects also use projects. In order to complete these sample projects, students need to use writing, research, and problem-solving skills they've learned in other classes:

- Students must complete an oral history based on interviews and written sources, and present their findings on tape as well as orally in class. The choice of subject matter is up to them. They create three workable hypotheses based on preliminary investigations, and four questions to test out each hypothesis.
- In science, the teacher issues a challenge and the students get together in the laboratory to solve it. They not only must come up with the answer, but keep a log and explain how they arrived at it.
- An English teacher assigns each student to write a poem and then join with other students to review and revise each one.
- Students in a high school business class might simulate a real-life situation, such as coming up with a new consumer product and devising a marketing plan to sell it. Frequently, students will have to go out in the community to do research for these projects.

Portfolios in Vermont

Vermont launched the first statewide portfolio assessment program in 1991-92. The program, piloted at sample sites the preceding year, required students in grades 4 and 8 to complete portfolios in writing and mathematics.

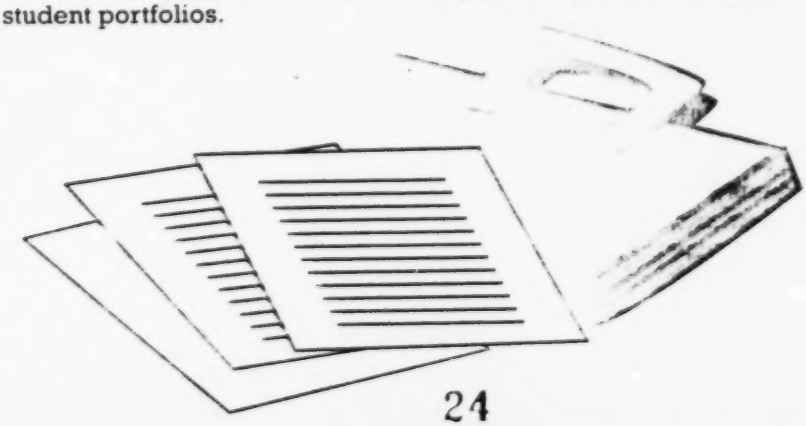
Writing portfolios contained examples of student work from the following four categories:

- A poem, play, or personal narration.
- A "personal response" to a cultural or sports event, book, math problem, or current issue.
- Prose pieces from classes other than English and language arts.
- A piece that students consider their "best."

Student work included in mathematics portfolios also had to meet problem-solving requirements, such as understanding what is being asked, selecting the right procedure for solving the problem, analyzing and verifying results, and looking beyond the answer at practical applications. In addition, responses were scored on the student's ability to use math language, charts, and graphs correctly, as well as on the overall clarity of the presentation.

"Uniform tests"—a combination of multiple-choice and open-ended questions—given to fourth and eighth graders complement the portfolio assessments.

Discussions are under way about expanding portfolio assessment to other subject areas. In the meantime, teachers, parents, and other educators continue learning about how to use and grade student portfolios.



Logic counts. In all these instances, the "solution" is only a part of the "answer." The students also must describe their assumptions, the processes they followed, and the reasons for their conclusions. In fact, even if the answer is not the expected one, they may still get some credit by showing their understanding of the concepts involved.

Portfolios

These are selected collections of student papers, drawings, essays, and other work that are collected by the teacher and reviewed periodically to determine progress. Portfolios often include a piece that the students consider their "best work" at the time, as well as some works in progress and other completed work that illustrate how the student has improved over time.

As opposed to traditional teachers' "folders" of student work, portfolios must meet assembly requirements and uniform evaluation procedures to be considered valid elements of an assessment program.

A Long Road Ahead

Obviously, the task of grading alternative assessments is far more complex—and expensive—than checking off correct answers or running an answer sheet through an optical scanner. An elaborate set of criteria must be developed by which to judge students' work. Teachers and other staff members need training in how to rate these performances objectively.

Many educators feel that teaching has gotten way ahead of testing, but performance-based assessments are working in many parts of the country to close this gap.

As new methods of testing and assessment continue to evolve, educators are finding ways to combine the strengths of each type of test in their annual testing programs. True standardization is difficult to achieve in performance assessments, so a limited use of norm-referenced standardized testing is often added to the mix. In addition, the publishers of standardized tests have begun to include some open-ended questions in their tests. Even the venerable SAT is adding some open-ended questions to its traditional multiple-choice format.

Conclusion

Parents may easily feel overwhelmed by the many types of assessments school systems give and by what these tests actually mean for their children.

Making sense of testing and assessment requires a commitment from parents and schools to work together. Schools need to stay abreast of the latest research and assessment developments and, in turn, inform parents in an understandable way. Parents need to ask questions and accept their responsibility for making sure their children take tests seriously. Both groups can learn from each other to help improve education for the most important group in society, our children.

Resources

The following organizations offer helpful publications on testing issues:

National PTA
700 North Rush Street
Chicago, IL 60611
(312) 787-0977

Educational Testing Service
Information Services
Princeton, NJ 08541-0001
(609) 921-9000

FairTest
342 Broadway
Cambridge, MA 02139
(617) 864-4810

Glossary

Accountability. The demand by public officials and taxpayers that school officials show how their investment in education is paying off. This concern has led to an increase in "accountability testing," which samples what a large number of students have learned and compares these findings with other schools and school systems. Accountability testing is in contrast to "instructional testing," where the main purpose is improving student learning in the classroom.

Achievement test. A test designed to measure school-taught learning. It usually covers basic skills, such as reading, language arts, and mathematics, but may cover other subjects as well.

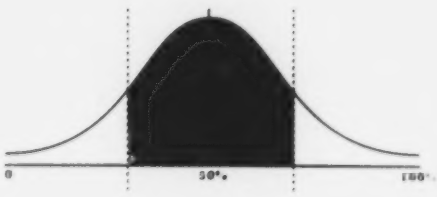
Alternative assessment. Any kind of assessment technique other than traditional testing methods, such as multiple-choice tests.

Aptitude test. A test designed to measure or predict a student's potential before instruction begins.

Assessment. An umbrella term for the process of evaluating students. In one sense, assessment is synonymous with "testing," but since today's assessments use techniques other than traditional "tests," including classroom observation, many educators prefer the larger term.

Authentic assessment. A general term for a method of alternative assessment that tests students' ability to solve problems or perform tasks in simulated "real-life" situations.

Bell curve. Also called a "normal curve," it is a symmetrical pattern for plotting the scores of a norm group so that exactly half of the scores fall above the midpoint and half below. All other scores are plotted within the curve to determine standardized test rankings.



Classroom test. Any test prepared by a teacher for use in the classroom. Since it is not standardized, the scores bear little or no relation to grades obtained by students on tests given in other classrooms.

Competency test. A test intended to make sure a student has met certain minimum standards of skills and knowledge. These tests are often used to determine whether a student should be promoted or graduated.

Glossary

Criterion-referenced test. A test with questions based on what the student was taught. It is designed to measure how much specific knowledge the student learned from that instruction.

Curriculum. The body of knowledge, skills, and attitudes to be transmitted to students. It includes such elements as what and how children are taught and what they are supposed to know and be able to do at the end of each curriculum unit.

Curriculum alignment. Making sure the curriculum and the tests match, ensuring that teachers cover material on the test, and that the content of the curriculum is tested.

Essay test. A test that requires students to answer questions in writing, emphasizing their recall, understanding, and reasoning rather than asking them to choose a correct answer.

Grade equivalent. One method of reporting a standardized test result. It represents the grade level at which a student performs. A score of 5.5 means that the student is doing as well as the "average" student halfway through the fifth grade.

Higher-order thinking skills. The ability of students to think creatively and solve problems.

High-stakes testing. Any testing program whose results have unusually important consequences for students, schools, school staff, or school systems. A testing program initiated under community pressure for accountability purposes could be considered high-stakes testing.

I.Q. (Intelligence Quotient) tests. These standardized norm-referenced tests were developed more than a century ago to measure a person's native intelligence. Because many educators question their reliability, they are not used as often today.

Multiple-choice test. A test consisting of questions or incomplete statements. Students must select the correct answer from the choices provided, sometimes by filling in a "bubble" on an answer sheet with a pencil.

Norm. The distribution of test scores obtained from the norm group (see below). The norm, by definition, is the midpoint of the performance of students in the norm group: 50 percent of them score above the norm and 50 percent below.

Glossary

Norm group. A random group of students that originally take a test under the authority of the test developer. Their performance establishes the baseline against which all other scores are compared.

Norm-referenced test. A test that relates the scores of each student to those of students in a control, or norm, group (see above). This test shows how each student and group of students rank compared with an average. The intention is that when schools and school districts give the same tests under the same conditions and are ranked according to the same norm, their scores will be comparable.

Objective. A goal for students to reach at the conclusion of a certain level of instruction. Objectives are determined by curriculum specialists in advance. An assessment program using criterion-referenced tests measures how well students have achieved each objective.

Percentile. This is another way of reporting test results. Percentiles range from 1 to 99, with 50 indicating the average. Rankings are determined by comparing each student's performance on the test with that of the norm group. If a student performed as well as or better on the test than 65 percent of the norm group, he or she is said to have scored at the 65th percentile.

Performance assessment. This type of alternative assessment, also called performance-based, requires students to perform a task to demonstrate knowledge or skills rather than just answer questions. Students may have to conduct a science experiment or explain in writing how they solved a math or science problem.

Portfolio. A file of student work used as an assessment technique. Specified contents of the portfolio are reviewed to determine the level of student performance and progress.

Project. A type of performance assessment. It uses a complex assignment that demands more than one type of activity from the student and a variety of means of reporting its conclusion.

Quartile. After percentiles are determined, the distribution may be broken down for reporting purposes into four equal groups: the 1-25th percentile, 26-50th, 51-75th, and 76-100th. Thus, a student who scored at the 45th percentile is said to be in the second quartile.

Glossary

Quintile. Similarly, percentiles sometimes are broken down into five groups: 1-20th, 21-40th, 41-60th, 61-80th, and 81-100th. Thus, the student in the 45th percentile is in the third quintile.

Reliability. A measure of consistency for tests. A reliable test will yield similar scores when the students' abilities or knowledge are similar.

Sampling. A way to get information about a large group by examining only a small but random number of the group (called the sample). When conducted properly, the results are considered highly reliable. This is how major public opinion polls are conducted, as are the federally funded tests called the National Assessment of Educational Progress.

Standard. An important assessment concept. It is the level of performance established by curriculum or test developers against which student performance is measured. The scores that determine the levels at which students "pass" or "fail," or are labeled "satisfactory" or "excellent," are arbitrarily established according to the judgment of the school system or state education department.

Stanine. Like percentile ranks, stanines show a student's standing relative to a norm group. Stanines range from a low of 1 to a high of 9, with 5 designating average performance.

Standardized test. An assessment instrument given to a large number of persons under similar conditions and designed to produce comparable scores. The term is most often applied to national norm-referenced tests developed by test publishers.

Teaching to the test. A questionable practice whereby teachers or schools concentrate instruction on topics they know will be on a test, drilling the students on types of information usually covered, solely to increase average test scores. This practice actually can raise student scores without improving student learning.

Test-taking skills. Techniques that can help students raise their scores. Many schools have introduced courses in test-taking skills, without revealing a test's content, to help students become more comfortable with the test format.

Validity. The measure of accuracy for tests. A test is considered valid when it measures what it is supposed to measure rather than extraneous variables.

Acknowledgments

Making Sense of Testing and Assessment was written to help educators explain, and parents and others understand, the myriad types of assessments students are given.

This booklet was written by Donald L. Hymes, who manages AASA's Critical Issues series and cowrote *The Changing Face of Testing and Assessment.* A longtime education writer, Hymes served as director of publications for the Montgomery County, Maryland, Public Schools.

Special thanks to AASA's Marion Hunt and to Ramsay Seldon, director of the State Education Assessment Center for the Council of Chief State School Officers, for reviewing the manuscript.

Gary Marx, AASA senior associate executive director, served as project director. Leslie Eckard, AASA publications manager, and Katie Ross, AASA communications assistant, edited the manuscript and supervised its production.

Graphic design was provided by Lea Croteau Bartlett of East Woods Studio in Sherwood Forest, Maryland.

1801 North Moore Street
Arlington, VA 22209-9988
(703) 875-0748